DOCUMENT RESUME

ED 191 907 TM 800 531

AUTHOR Gillmore, Gerry; Dickerson, Alice
TITLE The Relationship between Instruments Used for Identifying Children of Limited English Speaking Ability in Texas.
INSTITUTION Education Service Center Region 4, Houston, Tex.
SPONS AGENCY Texas Education Agency, Austin.
PUB DATE Nov 79
NOTE 37p.

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Achievement Tests; Classification; Correlation; Elementary Secondary Education; *English (Second Language); *Language Proficiency; *Language Tests; Second Language Learning; *Test Validity
IDENTIFIERS Basic Inventory of Natural Language; Bilingual Syntax Measure; Language Assessment Scales (DeAvila and Duncan); *Limited English Speaking; Primary Acquisition of Language; Primary Language Indicator Test (Schutt); Texas

ABSTRACT

Five language assessment tests (LATs) are used for identifying children of limited English speaking ability in Texas: Basic Inventory of Natural Language; Bilingual Syntax Measure; Language Assessment Scales; Primary Acquisition of Language; and Schutt Primary Language Indicator Test. Interrelationships between LATs and relationships between the LATs and achievement tests were studied. LATs were administered to a stratified random sample of 464 students, grades K-12, who had been previously identified as having been exposed to a language other than English. The subjects were randomly assigned to one of ten possible test pair groups of the five LATs. Correlation coefficients, Chi-Square values, and estimated communalities to indicate minimum reliabilities were calculated. The analyses indicated a lack of substantial relationship between the various LATs, and disagreement between the classification of English proficiency levels and achievement test performance. All evidence indicates a critical need for further clarification of classification of children of limited English proficiency and exit criteria from special programs. (Author/RL)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED191907

U S DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

T. Pate

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

THE RELATIONSHIP BETWEEN INSTRUMENTS USED FOR IDENTIFYING CHILDREN OF LIMITED ENGLISH SPEAKING ABILITY IN TEXAS

NOVEMBER 1979

THIS RESEARCH WAS CONDUCTED BY REGION IV EDUCATION SERVICE CENTER UNDER A TITLE VII GRANT FROM THE TEXAS EDUCATION AGENCY, CONTRACT #99675904.

BY

GERRY GILLMORE

ALICE DICKERSON

Region IV Education Service Center
Tom Pate, Jr., Executive Director
1750 Seamist
Houston, Texas 77008

A B S T R A C T

This project was undertaken to study the relationships between various tests used to assess students exposed to languages other than English. During the Spring, 1979 tests measuring English language proficiency were administered to 464 students, grades K-12, from six school districts in the greater Houston area; available achievement test scores were also collected.

The statistical analyses of the test data included the calculations of correlation coefficients, Chi Square values and estimated communalities to indicate minimum reliabilities. The analyses indicate a lack of substantial relationship between the various language assessment instruments. This may be due, in part, to lack of reliability of these tests, as well as dissimilarity of the aspects of English proficiency the tests purport to measure. There is disagreement between the classification of English proficiency levels and achievement test performance.

All evidence observed in this project indicates a critical need for further research, review, and clarification of classification of children of limited English proficiency and exit criteria from special program(s).

TABLE OF CONTENTS

		Page
ABSTRACT		ii
LIST OF TABLES		iv
	INTRODUCTION	1
	METHOD	3
	RESULTS	9
	The Relationship of the Five Language Assessment Tests	10
	The Relationship of the Language Assessment Instruments and Achievement Test Scores	14
	Special Analysis of the LAS Scores	15
	Special Analysis of the BINL Scores	18
	Reliability Estimates of Instruments	21
CONCLUSION		23
APPENDIX		
A	TEST DATA SUMMARY	25
B	CHI SQUARE VALUES AND PROBABILITIES	29
REFERENCES		30

LIST OF TABLES

Table		Page
1	LANGUAGE ASSESSMENT TESTS AND SCORES FOR DETERMINATION OF ENGLISH PROFICIENCY.....	2
2	NUMBER OF STUDENTS BY GRADE LEVEL AND PER "TEST-PAIR".....	4
3	ACHIEVEMENT TEST INFORMATION AVAILABLE.....	5
4	NUMBER OF SCORES BY TEST AND GRADE LEVEL.....	9
5	NUMBER OF STUDENTS WITH VARIOUS PAIRS OF TEST SCORES.....	10
6	RELATIONSHIPS OF THE FIVE LANGUAGE ASSESSMENT INSTRUMENTS FOR ALL STUDENTS COMBINED.....	11
7	RELATIONSHIPS OF THE FIVE LANGUAGE ASSESSMENT INSTRUMENTS BY GRADES.....	12
8	LESA AND NON-LESA IDENTIFICATION BASED ON FIVE LANGUAGE ASSESSMENT INSTRUMENTS.....	13
9	CLASSIFICATION OF STUDENTS AS LESA AND NON-LESA BY TEST PAIR.....	14
10	RELATIONSHIPS OF LANGUAGE ASSESSMENT AND ACHIEVEMENT TEST SCORES.....	15
11	LESA/NON-LESA CLASSIFICATION VERSUS ACHIEVEMENT TEST 23rd PERCENTILE EXIT CRITERION.....	16
12	LESA/NON-LESA CLASSIFICATION VERSUS ACHIEVEMENT TEST 40th PERCENTILE EXIT CRITERION.....	16
13	RELATIONSHIPS OF LAS, OTHER LANGUAGE ASSESSMENT AND ACHIEVEMENT TEST SCORES.....	17
14	RELATIONSHIPS OF BINL-B, C, & D SCORES BY GRADES WITH OTHER LANGUAGE ASSESSMENT INSTRUMENTS.....	19

Table		Page
15	THE CLASSIFICATION OF STUDENTS AS LESA AND NON-LESA BY TEST PAIR USING BINL-B, C, & D.....	19
16	RELATIONSHIPS OF BINL-B, C, & D SCORES WITH ACHIEVEMENT SUBTEST SCORES.....	20
17	LESA/NON-LESA BINL CLASSIFICATION VERSUS ACHIEVEMENT TEST EXIT CRITERION.....	20
18	RELIABILITY ESTIMATES OF TEST INSTRUMENTS.....	22

THE RELATIONSHIP BETWEEN INSTRUMENTS USED FOR IDENTIFYING CHILDREN OF LIMITED ENGLISH SPEAKING ABILITY IN TEXAS

INTRODUCTION

The *Lau v. Nichols* decision prompted the Office of Civil Rights to establish a set of remedies for meeting the educational needs of LESA (limited English speaking ability) children. Possible LESA children are those with any exposure to a language other than English. While parental and teacher opinion may be considered, the Office of Civil Rights requires that some measure be administered to all other language exposed children to determine English speaking ability. In order to respond to this requirement, the Texas Education Agency identified five instruments which could be utilized by school districts to measure English speaking ability.

A survey of the test manuals and materials indicates that each of the instruments is based on a different theoretical approach; each provides varying supportive research information. A descriptive summary of the reported test data supplied for the English version of each instrument may be found in Appendix A; the scores on the English version which classify children as LESA are indicated in Table 1. The scores are available as levels and/or raw scores. Three instruments use a 1 to 5 scale to determine English proficiency; one uses a minimum raw score and a 1 to 4 scale for determining level; another uses a minimum raw score in two areas, listening and fluency, which can be converted to standard scores provided by the test developer.

Very little information was found in a review of the literature. A linguistic analysis of some of the instruments indicates that they may "fail to measure what they are intended to" (Dietrich, Freeman and

TABLE 1

LANGUAGE ASSESSMENT TESTS AND SCORES

FOR DETERMINATION OF ENGLISH PROFICIENCY

TEST	GRADE LEVEL	MINIMUM SCORE REQUIRED FOR NON/LESA CLASSIFICATION	
		RAW SCORE	LEVEL
BINL (Basic Inventory of Natural Language)	K-2	50	4
	3-8	75	4
	9-12	100	4
BSM (Bilingual Syntax Measure)			
Level I	K-2		4
Level II	3-12		5
LAS (Language Assessment Scales)			
Level I	K-5	75	4
Level II	6-12	82	4
PAL (Primary Acquisition of Language)			
OLDM (Oral Language Dominance Measure)	K		4.5
	1-3		5
OLPM (Oral Language Proficiency Measure)	4-6		5
SPLIT (Schutt Primary Language Indicator Test)		LISTENING	FLUENCY*
	K	10	6
	1	14	7
	2-3	15	9
	4	16	9
	5-6	17	9

*Appropriate standard scores were used for analyses.

Crandall, 1979). While other studies have examined exit criterion from bilingual programs (De Avila and Duncan, 1978), no information is presently available as to the commonality for placement of children into the same program. Since the five instruments are being used to determine English speaking ability, since they all purport to measure the same thing using different theoretical bases, and since technical information varies

tremendously, it is of value to determine the extent of the relationship which exists between any two of the instruments. A tenable hypothesis is that two or more tests administered to the same child would identify the child's English speaking ability at the same level.

Following the Fall, 1978 administration of the testing program designed to identify LESA children, several school districts within Region IV raised questions regarding the application of the language assessment instruments. After discussion with representatives of the Texas Education Agency, the test publishers and the school districts, a research project was implemented. The study includes the interrelationships of the five language assessment instruments and the criteria for exit from the instructional programs for LESA children.

METHOD

The study was conducted in six school districts in the greater Houston area (Alief, Alvin, Dickinson, Hitchcock, Sheldon, Wallis-Orchard) with test data collected during March and April, 1979. Children in grades K-12 who were exposed to a language other than English had been previously identified using the procedure outlined by the Texas Education Agency in the "Texas State Plan for Bilingual Education."* With the exception of one district, all such identified students were previously administered one of the five language assessment instruments during the Fall of 1978.

The 464 participants in this study were selected in such a way as to result in a stratified random sample from the described population of 2,523 bilingual students. According to the grade level range of applicability of the instruments, the subjects were randomly assigned to one

*Texas Education Agency, "Texas State Plan for Bilingual Education," November 11, 1978.

of the ten possible test pair groups of the five language assessment instruments so as to obtain adequate data per test per grade level for valid statistical procedures. The number of students whose test scores were analyzed is shown in Table 2. Each participant's original language

TABLE 2

NUMBER OF STUDENTS BY GRADE LEVEL

AND PER "TEST-PAIR"

GRADE	N	BINL BSM	BINL LAS	BINL PAL	BINL SPLIT	BSM LAS	BSM PAL	BSM SPLIT	LAS PAL	LAS SPLIT	PAL SPLIT
K	51	11	9	7	8	7	5	3	7	4	7
1	52	10	6	9	7	9	6	6	7	7	8
2	57	12	10	13	8	8	8	6	7	6	8
3	49	12	12	11	8	7	6	7	8	6	9
4	51	8	10	9	7	7	6	5	9	6	9
5	56	10	10	10	12	5	7	8	9	8	8
6	48	10	12	8	11	8	10	9	10	8	9
7	18	12	11			13					
8	19	11	10			15					
9	18	5	6			17					
10	15	5	7			13					
11	17	6	7			16					
12	13	4	4			9					
K-12 TOTAL	464	116	114	67	61	134	48	44	57	45	44

N. B. PAL and SPLIT are not appropriate for grades 7-12.

assessment score was obtained for analysis or, in the case of one district, a third instrument was administered so that three of the five test scores were available for determination of each child's English language proficiency.

Since the Texas Education Agency specifies national percentile rank on the reading and language sections of approved achievement tests as exit criteria, the name of the achievement test and the national percentile scores on the appropriate subtests were obtained when available. Table 3 identifies the number of available scores for the sample.

TABLE 3

ACHIEVEMENT TEST INFORMATION AVAILABLE

TEST	N
California Achievement Tests, 1977 Edition	21
Comprehensive Test of Basic Skills	17
Iowa Test of Basic Skills	49
Metropolitan Achievement Test	2
Sequential Test of Educational Progress	11
SRA Achievement Series	64
Stanford Achievement Test	34
Iowa Test of Educational Development	5
	203

University students and an individual previously used in testing programs were trained in the administration of the five language assessment instruments. Since the Texas Education Agency had identified contact persons for each instrument, these individuals were responsible

for conducting the sessions training test administrators in the use of their instruments. All individuals were trained in the administration and scoring of the five instruments.

The statistic chosen to measure the relationship of the various language assessment instruments is Kendall's Tau, a non-parametric correlation coefficient. This procedure does not require a normal distribution of the scores or more than an ordinal level of measurement. The relationship of the language assessment instruments and the various achievement test scores is measured by the Pearson product-moment correlation coefficient. There is sufficient range of these scores to justify the use of this more traditional statistic. The significance of the correlation coefficient is derived from the use of *Student's* t with N - 2 degrees of freedom. The one-tailed test of significance is used since all the measures of student performance are ordered the same (high to low). Values with probabilities of less than .05 are considered significant.

Other kinds of relationships are presented in tabular form. The statistical significance of these relationships can be measured by a variety of procedures, including Chi Square. These values and probabilities are reported where applicable in Appendix B. However, the real issue is the educational implications of these relationships. Therefore, the conclusions and recommendations will be based on a combination of statistical and educational significance.

The current Texas Education Agency established criterion for exit from a bilingual program is attainment of at least the 23rd percentile on both the reading and language section of a standardized achievement test. The previous criterion was the 40th percentile. Both of these

values were utilized to establish an exit and non-exit evaluation of the student. The scores of four of the language assessment instruments were expressed as a level which indicates various language proficiencies. These levels were obtained by adherence to the test manual specifications.

In the case of the SPLIT, the test representative who conducted the administration and scoring sessions stated that it would be impossible to calculate a level since the test was to be administered in English only. Thus, the SPLIT analysis utilized a score which was obtained by transforming raw scores for Listening and Verbal Fluency (see Table 1) to standard scores according to tables published for use with the SPLIT; the two standard scores were added to provide a raw score for analysis purposes.

Each of the language assessment instruments has an established basis on which a child is classified as LESA or non-LESA. The cutoff points which were utilized to calculate this classification for each student are specified in Table 1.

Two of the five tests were included in additional special analyses due to questions raised during the discussions with test and school district representatives; these two tests are the BINL and the LAS. Special treatment of the LAS occurred at the request of one of the test representatives; he theorized that Part V (Storytelling) of this instrument parallels some of the other instruments more than the total LAS test. If true, the subscore of LAS, Part V, should have a higher correlation than the total LAS score with some or all of the other instruments.

Additional treatment of the BINL was implemented to investigate the variations in scoring procedures for tests submitted to and returned by the publisher and scorer, CHECpoint Systems, Inc. When school

districts submitted student responses on the BINL to CHECpoint Systems, Inc. for computer scoring during the Fall, 1978, the BINL scores reported were labeled "Pre-Test English B." During the next several months, it was discovered that other data sent for scoring were returned labeled "Pre-Test English C" and "Pre-Test English D."

Verbal and written communications were initiated with Dr. Charles Herbert of CHECpoint, Systems, Inc. in an effort to understand the changes in the scores. A summary of the communications resulted in these salient points:

1. The changes in the three scoring procedures involve the "consideration of verb auxiliaries and adjective placement as well as consideration of the pattern type for each sentence."
2. The result of the changes in the scoring procedures is that "the general trend of the three sets of scores...is upward."
3. Research is being conducted by CHECpoint to "further refine our minimum/maximum ranges by grade levels." (These are the bases for classification as a LESA student by the BINL test.)

CHECpoint rescored the original student responses for the participating districts and furnished those results to the districts which resulted in available "B", "C" and "D" scores. The intent of the inclusion of the three BINL scores for analysis is to determine if the changes did, in fact, result in more accurate classification of the students as LESA or non-LESA individuals.

Since a district submitting BINL responses to CHECpoint for scoring

would have no choice as to the scoring procedures used, the majority of the comparisons of the five language assessment instruments utilize the scores currently being reported, BINL-D.

RESULTS

The number of student scores available for analyses is shown in Table 4 and the number per test pairs including achievement test scores is

TABLE 4

NUMBER OF SCORES BY TEST AND GRADE LEVEL

GRADE LEVEL	TESTS						
					BINL		
	BSM	LAS	PAL	SPLIT	B	C	D
K	21	20	20	18	11	22	22
1	24	24	24	22	11	18	18
2	25	25	23	23	13	25	25
3	24	23	23	21	13	23	23
4	22	23	23	23	11	19	18
5	23	23	24	25	8	23	23
6	23	24	24	22	13	22	22
7	15	15			9	13	12
8	17	17			12	12	10
9	17	18			8	6	6
10	13	15			4	7	7
11	16	17			4	7	7
12	11	9			3	5	5
TOTAL K-12	251	253	161	154	120	202	198

N. B. PAL and SPLIT are not appropriate for grades 7-12.

indicated in Table 5. These N's, applicable to the statistical results, are omitted in the relationship sections for clarity of presentation.

TABLE 5

NUMBER OF STUDENTS WITH

VARIOUS PAIRS OF TEST SCORES

	BSM	LAS	SPLIT	PAL	BINL B	BINL C	BINL D
LAS	134						
SPLIT	44	46					
PAL	48	58	58				
BINL-B	78	79	34	39			
BINL-C	116	116	62	68	114		
BINL-D	114	114	61	67	113	198	
VOCABULARY	102	97	67	64	24	49	47
COMPREHENSION	106	101	66	63	28	53	51
READING	98	96	64	62	11	42	40
LANGUAGE	100	97	63	65	24	55	53
MATHEMATICS	108	106	71	68	28	59	57

The Relationship of the Five Language Assessment Tests

The intercorrelation matrix of the five language assessment instruments is presented in Table 6 for all students combined. In every instance, the correlation between the BSM and the other language assessment measures was the highest. The instruments showed relationships with each other that are

TABLE 6

RELATIONSHIPS OF THE FIVE LANGUAGE ASSESSMENT INSTRUMENTS FOR ALL STUDENTS COMBINED

	BINL	BSM	LAS	PAL
BSM	.31*			
LAS	.24*	.38*		
PAL	.24*	.33*	.33*	
SPLIT	.09	.42*	.24*	.34*

*Values significant at .05 level. Correlation coefficients are Kendall's Tau.

statistically significant with the exception of the BINL and the SPLIT. However, the correlation coefficients (.24 to .42) indicate that there is only a slight to moderate relationship in the ways any two of these tests are measuring language ability.

The correlation coefficients of the test-pairs by grade spans are shown in Table 7. Across grade level spans the BSM is the most stable co-varying measure except when paired with the SPLIT. The greatest variance of the correlation coefficients occurs between the SPLIT and the other tests. The only significant correlations observed involving the SPLIT are the SPLIT-BINL at the K-2 level, and the SPLIT-PAL at the 3-6 level.

The highest correlation observed at the K-2 level is between the BINL-LAS; at the 3-6 level, the BSM-LAS; over all grades, the BSM-SPLIT. The test pair showing the most relationship at all grades and individual grade spans is the BSM-LAS.

TABLE 7

RELATIONSHIPS OF THE FIVE LANGUAGE ASSESSMENT INSTRUMENTS BY GRADES

	K-2	3-6	7-12	K-12
BINL/BSM	.37*	.24*	.31*	.31*
BINL/LAS	.48*	.27*	.07	.24*
BINL/PAL	.31*	.21*		.24*
BINL/SPLIT	.26*	-.11		.09
BSM/LAS	.42*	.52*	.32*	.38*
BSM/PAL	.23	.34*		.33*
BSM/SPLIT	-.18	.21		.42*
LAS/PAL	.34*	.30*		.33*
LAS/SPLIT	.08	-.04		.24*
PAL/SPLIT	.21	.25*		.34*

*Values significant at .05 level. Correlation coefficients are Kendall's Tau.
N. B. PAL and SPLIT are not appropriate for grades 7-12.

The five language assessment instruments were also analyzed as to their classification of LESA children. Results presented in Table 8 indicate the number and percentage of children's LESA classification according to individual tests. Additionally, the scores of students with three assessments were summarized according to agreement of LESA classifications. Of the 183 students with three language assessment scores, 58 (32%) had three agreeing LESA classifications and 148 (68%) had disagreeing LESA classifications.

TABLE 8

LESA AND NON-LESA IDENTIFICATION BASED ON

FIVE LANGUAGE ASSESSMENT INSTRUMENTS

	NUMBER OF STUDENTS		PERCENTAGE	
TEST	NON LESA	LESA	NON LESA	LESA
BINL	54	144	27%	73%
BSM	203	48	81%	19%
LAS	113	49	70%	30%
PAL	60	101	37%	63%
SPLIT	153	1	99%	1%

The PAL and the BINL identify more LESA students than do the other three tests. Because the SPLIT classification resulted in only one student being identified as LESA, further comparisons involving the SPLIT are not included.

Table 9 contains the classifications of students based on pairs of language assessment instruments. Any two of the four language assessment instruments appear to agree on the LESA classification of a child from 45%-78% of the time. The BSM and the LAS show the highest agreement; this is consistent with the relationship measured by correlation coefficients.

TABLE 9

CLASSIFICATION OF STUDENTS AS LESA AND NON-LESA

BY TEST PAIR

TEST PAIR	CLASSIFICATION			PERCENTAGE	
	AGREE		DISAGREE		
	NON-LESA	LESA		AGREE	DISAGREE
BINL/BSM	29	22	63	45%	55%
BINL/LAS	13	21	35	49%	51%
BINL/PAL	5	35	27	60%	40%
BSM/LAS	31	9	11	78%	22%
BSM/PAL	19	6	23	52%	48%
LAS/PAL	21	14	22	61%	39%
TOTAL	118	107	181	55%	46%

The Relationship of the Language Assessment Instruments and Achievement Test Scores

The correlation coefficients of the five language assessment tests and the achievement test scores are shown in Table 10. Two of the language tests have a significant correlation with all five achievement subtest scores. The correlations between the PAL and the achievement test scores are the highest observed r's, with the LAS yielding the next highest r's. The BINL, BSM, and SPLIT show no significant relationship with any of the achievement test scores.

The question of the LESA and non-LESA identification of a student versus the achievement test scores for exit criteria was also studied. This was done using the Texas Education Agency established levels for

TABLE 10

RELATIONSHIPS OF LANGUAGE ASSESSMENT

AND ACHIEVEMENT TEST SCORES

	ACHIEVEMENT SUBTEST SCORES				
TEST	VOCAB-ULARY	COMPRE-HENSION	READ-ING	LAN-GUAGE	MATHE-MATICS
BINL	-.24	-.07	.02	-.21	-.19
BSM	.02	.15	-.03	.12	.01
LAS	.28*	.33*	.31*	.38*	.33*
PAL	.42*	.33*	.46*	.41*	.41*
SPLIT	.11	.07	.08	.11	.14

*Values significant at .05 level. Correlation coefficients are Pearson's *r*.

LESA classification at the 23rd percentile requirement for exit. The old percentile requirement was also considered. Tables 11 and 12 present these findings.

The data contained in Tables 11 and 12 indicate that the changing of the exit criterion from the 40th to the 23rd percentile on the reading and language achievement test scores resulted in more agreement between the two classifications. Using the 23rd percentile exit criterion, one-third to one half of the language assessment classifications disagreed with the achievement score exit assessment.

Special Analysis of the LAS scores.

The LAS test scores were analyzed using the following:

1. LAS raw score of subtests 1-4
2. LAS raw score of subtest 5

TABLE 11

LESA/NON-LESA CLASSIFICATION VERSUS ACHIEVEMENT TEST

23rd PERCENTILE EXIT CRITERION

	CLASSIFICATION					
	EXIT		NON-EXIT		PERCENTAGE	
TEST	NON-LESA	LESA	NON-LESA	LESA	AGREE	DISAGREE
BINL	9	7	11	13	55%	45%
BSM	40	8	43	2	45%	55%
LAS	42	5	26	16	65%	35%
PAL	21	15	7	16	63%	37%
SPLIT	39	0	18	0	68%	32%
TOTAL	151	35	105	47	59%	41%
Percentage	41%	10%	71%	14%		

TABLE 12

LESA/NON-LESA CLASSIFICATION VERSUS ACHIEVEMENT TEST

40th PERCENTILE EXIT CRITERION

	CLASSIFICATION					
	EXIT		NON-EXIT		PERCENTAGE	
TEST	NON-LESA	LESA	NON-LESA	LESA	AGREE	DISAGREE
BINL	3	4	17	16	48%	52%
BSM	23	4	60	6	31%	69%
LAS	28	3	40	18	52%	48%
PAL	13	4	15	27	68%	32%
SPLIT	24	0	33	0	42%	58%
TOTAL	91	15	165	67	47%	53%
Percentage	27%	4%	49%	20%		

3. LAS total raw score

4. LAS level.

Table 13 presents the relationships of the various LAS, language assessment and achievement test scores. The correlation coefficients involving the subtests 1-4 raw scores are higher than those of the other language assessment instruments and the other LAS scores. For all five achievement subtest scores, the LAS score which yields the highest correlation coefficient is the LAS level. This is the language assessment index used in the other analyses.

TABLE 13

RELATIONSHIPS OF LAS, OTHER LANGUAGE ASSESSMENT AND ACHIEVEMENT TEST SCORES

	LAS			
TEST	1-4	5	RS	LEVEL
BINL	.32*	.23*	.26*	.24*
BSM	.46*	.39*	.44*	.38*
PAL	.36*	.19*	.26*	.33*
SPLIT	.31*	.04	.14	.24*
VOCABULARY	.14	.04	.07	.28*
COMPREHENSIVE	.24*	.14	.19*	.33*
READING	.14	.03	.07	.31*
LANGUAGE	.16	.25*	.27*	.38*
MATHEMATICS	.13	.21*	.22*	.33*

*Values significant at .05 level. Correlation coefficients of language assessment instruments are Kendall's Tau. Correlation coefficients of achievement test scores are Pearson's r.

The extended analyses of the LAS provide evidence to substantiate the use of the LAS level, as opposed to the subtest scores to measure relationships between other assessment and achievement measures. In fact, the lowest observed intercorrelations of the LAS and other language tests occurred using the LAS subtest 5 score, which was the score hypothesized to correlate highest with these other measures.

Special Analysis of the BINL Scores

Tables 14 and 15 depict the results of the analyses of the three BINL scores: BINL-B, BINL-C, and BINL-D and other language assessment instruments. The evidence in Table 14 indicates that the adjustment of the BINL scores has resulted in little, if any, increase in correlation coefficients. The data in Table 15 indicate that the relationship between the BINL classification and the BSM and LAS classifications improved with the adjustment of the BINL scores. The percentage of agreement decreased with the PAL.

The percentage of agreement of LESA classification based on pairs of the three BINL scores ranges from 94% to 77%. The adjustments, which changed some of the classifications, resulted in a higher percentage of agreement with two of the three other language assessment instruments.

The results shown in Table 16 indicate a lack of relationship between BINL language assessment scores and standardized achievement test scores. The negative values of the correlation coefficients indicate that the higher the achievement test score, the lower the English ability level; conversely, the lower the achievement test score, the higher the English ability level. The adjustments made to the scores do not increase the relationship.

TABLE 14

RELATIONSHIP OF BINL-B, C, & D SCORES BY GRADES

WITH OTHER LANGUAGE ASSESSMENT INSTRUMENTS

	K-2			3-6			7-12			K-12		
	BINL			BINL			BINL			BINL		
TESTS	B	C	D	B	C	D	B	C	D	B	C	D
BSM	.42*	.46*	.37*	.50*	.21*	.24*	.47*	.34*	.31*	.46*	.34*	.31*
LAS	.36*	.33*	.48*	.31*	.30*	.27*	.24*	.00	.07	.24*	.17*	.24*
PAL	.15	.23*	.31*	.23	.24*	.21*				.18*	.23*	.24*
SPLIT	.37*	.28*	.26*	-.02	-.18	-.11				.08	-.02	.09

*Values significant at .05 level. Correlation coefficients are Kendall's Tau.

N. B. PAL and SPLIT are not appropriate for grades 7-12.

TABLE 15

THE CLASSIFICATION OF STUDENTS AS LESA AND NON-LESA

BY TEST PAIR USING BINL-B, C, & D

	CLASSIFICATION BY OTHER INSTRUMENT					
	NON-LESA		LESA		PERCENTAGE	
	BINL CLASSIFICATION		BINL CLASSIFICATION			
TEST PAIR	NON-LESA	LESA	NON-LESA	LESA	AGREE	DISAGREE
BSM/BINL-B	4	60	0	14	23%	77%
BSM/BINL-C	12	78	1	25	32%	68%
BSM/BINL-D	29	59	4	22	45%	55%
LAS/BINL-B	1	27	0	11	31%	69%
LAS/BINL-C	4	42	0	23	39%	61%
LAS/BINL-D	13	33	2	21	49%	51%
PAL/BINL-B	0	11	0	28	72%	28%
PAL/BINL-C	1	22	4	41	62%	38%
PAL/BINL-D	5	18	9	35	60%	40%
BINL-B/BINL-C	4	0	7	103	94%	6%
BINL-B/BINL-D	4	0	26	83	77%	23%
BINL-C/BINL-D	16	0	38	144	81%	19%

TABLE 16

RELATIONSHIPS OF BINL-B, C, & D SCORES WITH

ACHIEVEMENT SUBTEST SCORES

SUBTESTS	BINL-B	BINL-C	BINL-D
Vocabulary	-.47*	-.26*	-.24
Comprehension	-.15	-.11	-.07
Reading	.34†	-.01	.02
Language	-.41*	-.11	-.21
Mathematics	-.34*	-.15	-.19

*Values significant at .05 level. Correlation coefficients are Pearson's *r*..
†Not significant, N=11.

The agreement in LESA classification and exit criterion is shown in Table 17 with the adjustment of the BINL scores. The percentage of agreement is greatest using BINL-B and lessens with each adjustment.

TABLE 17

LESA/NON-LESA BINL CLASSIFICATION VERSUS

ACHIEVEMENT TEST EXIT CRITERION*

BINL	CLASSIFICATION				PERCENTAGE	
	EXIT		NON EXIT			
	NON-LESA	LESA	NON-LESA	LESA	AGREE	DISAGREE
B	2	2	1	6	73%	27%
C	4	12	5	21	60%	40%
D	9	7	11	13	55%	45%

*Based on criterion of 23rd percentile.

The analyses of the BINL scores leads to questioning of the adjustments. Overall, the BINL-B scores appear to be as appropriate as either of the other adjusted scores. Regardless of the BINL scores used, there is no valid significant relationship with other instruments. In fact, the negative correlation coefficients with the achievement subtest scores indicating inverse relationships cannot be valid. Serious consideration needs to be given to these instruments if either or both is to be used to qualify for, or exit from, the same program.

Reliability Estimates of Instruments

Because of the low correlation coefficients obtained in the analyses of these data, one further statistical procedure was carried out. An estimate of the communalities was obtained to serve as an indicator of the minimum reliability of the various instruments. This was done based on the theory that perhaps the correlation coefficients were as low as they are because of the lack of reliability of the tests.

The estimated communalities, i.e., minimum reliabilities, are shown in Table 18. The values for the achievement test scores (.7 - .9) are comparable to the values usually reported for these type measurements. The values for the language assessment instruments (.2 - .5) indicate that the reliabilities of these instruments are lower than reported in the technical materials and, in fact, may be the explanation of the low relationship between any two of the instruments.

TABLE 18

RELIABILITY ESTIMATES OF

TEST INSTRUMENTS

Achievement Test Subscores	
Vocabulary	.94
Comprehension	.92
Reading	.98
Language	.68
Mathematics	.67
Language Assessment Instruments	
BINL	.22
BSM	.53
LAS	.32
PAL	.31
SPLIT	.43

N. B. Obtained by factor analytic procedure for estimated communalities yielding minimum values of reliability.

CONCLUSION

The sample included in this research was drawn from a population of six school districts located in the greater Houston area in Southeast Texas. Within the given time and resource constraints, the sampling procedures resulted in an adequate number of students within each possible cell as determined by test pair and grade level. The conclusions must be considered within the scope of the original population with generalizations to a greater population regarded cautiously.

The correlation coefficients of the various language assessment instruments furnish evidence that these instruments are measuring English ability in different ways. The lack of agreement is supported by the classification of LESA and non-LESA. The analyses that include consideration of grade levels show that some of the instruments vary widely across grade levels resulting in correlation coefficients that are higher than the true relationships. This occurs most frequently when comparisons involve the SPLIT. Since administration procedures required use of English scores alone and conversion to SPLIT standard scores rather than levels, the resulting measures of relationship are open to question.

While this project was to study the intercorrelations of the five language assessment instruments, analyses were expanded to include the required performance for exit from the programs for LESA children. Therefore, achievement test performance was considered. Only two of the five language assessment instruments, LAS and PAL, have statistically significant relationships with the achievement subtest scores. Observations of the Reading and Language subtests indicate that 10% of the children classified as LESA were qualified to exit the program. Therefore, at least 10% of the children are misclassified by one or both instruments.

The relationships among the language assessment instruments vary from nonexistent to moderate. Similar relationships were observed among the language assessment instruments and achievement tests. Since test theory states that an instrument cannot correlate higher with another instrument than its reliability, it is possible that the low relationships are due to the low reliabilities of the language assessment instruments.

All aspects of classification of LESA children and exit criteria from special programs must be reviewed. The fact that there are few instruments available that claim to measure English language proficiency is no reason to choose one of them for educational decisions. Political and/or legal judgements should not be the determining factors upon which to base appropriate educational programs. There must be immediate clarification through further scientific research to find educationally sound answers with regard to children who have language backgrounds other than English.

APPENDIX A

TEST DATA SUMMARY

Basic Inventory of Natural Language

The BINL obtains speech samples and attempts to eliminate inhibitors related to specific content, questioning, roles, situations and distractors. Although the test results are reported to include measures of language dominance, fluency, complexity of language and language development or growth, all results are based on the "Level of Complexity." Means, standard deviations and standard error are reported on five school districts from California and Illinois in urban and rural settings. Children in the sample participated in Elementary and Secondary Education Act Title I or Bilingual K-8 programs. Norm tables are provided to districts upon request. No further technical information as to validity of the measure or its reliability are furnished.

Bilingual Syntax Measure

The BSM was developed to assess language development as indicated by syntactic proficiency. Syntax is defined as the system of rules for arrangement, interrelationship and form of words, phrases and sentences in a language. Syntactical structure is measured on the assumption that a second language make errors similar to those who learn the language as a first language. For Level I, the field research was conducted in four geographic areas of the U. S. with a sample of 1,572 including seven ethnic groups.

Validity was established by using the SAI (an index reflecting the degree of grammaticality of a child's total speech corpus). Two scales were established: H-1 which consists of eight items eliciting structures

distinctive for a Level 5 proficiency, and H-2 which consists of ten items eliciting structures distinctive to Level 4 speakers. Validity was established by comparing each test item with the appropriate H-1 and H-2 score. The correlation between the H-1 scales and the SAI is .78 and between the H-2 scales and the SAI is .81.

Reliability of the instrument was determined by test-retest method and interscorer agreement was greater than 80%. Cohen's Kappa coefficient was used to exclude chance agreements between test administrations. The Kappas coefficients were $k = .45$, $k_{w1} = .51$ and $k_{w2} = .62$ with a standard error of .064. To determine language dominance the Kappa coefficients were $k = .46$, $k_{w1} = .54$ and $k_{w2} = .64$ with a standard error of .053. The authors stress that reliability coefficients reflect the variability of the acquisition process of language and should not be compared to those obtained for standardized tests. At the present time, the Technical Handbook is not available for Level II.

Language Assessment Scales

The LAS incorporates the phonemic, referential (lexical), syntactical and pragmatic systems to determine the linguistic abilities of children. For Level I, studies on interrator agreement ranged between .86 and .96. Validity was measured by use of tests to determine if English speakers performed significantly different from the LESA group; the differences between English speaking groups and Spanish speaking groups were significant at the .001 level. Internal reliability excluding syntax production was determined using the Cronbach Alpha coefficient and ranged in value from .68 to .95.

In preparing Level II, similar studies were conducted. Through the use of Analysis of Variance, comparisons were made between performances

of the English dominant and Spanish dominant children on the English form of the test; there were statistically significant differences for all subtests and total score ($p < 0.01$). Three separate studies were conducted to establish inter-judge reliability and the range of agreement was from .80 to 1.00. Interscale correlations between subtests range from .41 to .83.

Primary Acquisition of Language

A child's oral proficiency is based on the concept that proficiency requires four basic abilities: the ability to hear an utterance, understand it, formulate a statement and produce it. Scores are obtained by determining the number of acceptable responses and the total number of words in the responses. A level of proficiency is obtained by using a two-way table of the above two values. The table, norms and cutoff scores were based on field test data from 200 student responses representative of the school districts in the El Paso Schools.

The OLDM and OLPM were validated by ranking schools of predominant English speaking children from high socioeconomic levels to predominant Spanish speaking children from low socioeconomic levels. Concurrent validity was based on the expectation that there should be a positive correlation between the original ranking and the English mean score (a negative one was expected with the Spanish scores). Correlation coefficients were reported at .88 with the socioeconomic ranking correlating positively with the English means. Further concurrent validity data presented a positive correlation between English reading scores on the SRA and OLPM English scores (.54 to .45).

Item analysis data are reported with item difficulties for the OLDM ranging from .6 to .8 and the OLPM from .77 to .99. Internal consistency

for the English test using the Kuder-Richardson formula 20 is reported as r = .98 for the OLDM and as r = .92 for the OLPM.

Schutt Primary Language Indicator Test

The author specifies that "items and procedures of the SPLIT are culture bound" since language is a normal outgrowth of culture. Descriptive demographic data are presented on the standardization group (N = 1033) with regard to age, grade, sex, cultural heritage, residence, class rank and bilingualism. Appropriate subtests were administered to the standardization group and norm tables were calculated. These included the raw score converted to standard scores, the raw score converted to age and grade equivalents and raw scores converted to percentiles, by grade level for the six subtests.

The "Language Dominance Indicator" was calculated based on the total test score (standard score) difference between Spanish and English. The tables furnish cut-off scores for five dominance levels for two different grade level spans, K-2 and 3-6. Reliability was determined by using equivalent halves of the instrument and Kuder-Richarsdon formula 20. For the total test, reliability coefficients for 5 to 12 year olds ranged from .67 (12 years) to .80 (8 years). The reliability coefficients of the subtest scores were also presented and were in the .50 to .80 range.

Correlation between teacher judgments and raw scores in English (r = .70) provided one measure of validity. Further, a small sample (N = 16) of children were given the SPLIT and another language assessment measurement (the Pictorial Test of Bilingualism and Language Dominance); the correlation coefficients of the SPLIT and PTBLD ranged from .61 to .88 across grade levels.

APPENDIX B

CHI SQUARE VALUES AND PROBABILITIES

FOR

TABLE 9 RELATIONSHIPS

TEST PAIR	CHI SQUARE	P
BINL/BSM	2.22	>.13
BINL/LAS	2.40	>.12
BINL/PAL	.04	>.84
BSM/LAS	10.82	<.01
BSM/PAL	.04	>.84
LAS/PAL	4.60	<.05

CHI SQUARE VALUES AND PROBABILITIES

FOR

TABLE 11 AND 12 RELATIONSHIPS

TEST	23 PERCENTILE EXIT CRITERION		40 PERCENTILE EXIT CRITERION	
	CHI SQUARE	P	CHI SQUARE	P
BINL	.10	>.74	.0	>.99
BSM	2.45	>.11	0.19	>.65
LAS	7.81	<.01	4.00	<.05
PAL	3.33	>.06	6.51	<.03
SPLIT	NOT APPROPRIATE			

REFERENCES

Burt, M. K., Dulay, H. C. & Hernandez Ch., E. Bilingual Syntax Measure Technical Handbook. Oakland, CA: Harcourt Brace Jovanovich, 1975.

De Avila, E. A., Duncan, S. E. & Cervantes, R. A. Bilingual Program Exit Criteria. Report presented to the California State Department of Education, Office of Program Evaluation and Research, Sacramento, CA: May, 1978.

De Avila, E. A. & Duncan, S. E. Entry/Exit Criteria. Position paper prepared for the California State Department of Education, Sacramento, CA, August, 1978.

De Avila, E. A. & Duncan, S. E. Language Assessment Scales, Level I. Corte Madera, CA: Linguametrics Group, 1976.

De Avila, E. A. & Duncan, S. E. Language Assessment Scales, Level II. Corte Madera, CA: Linguametrics Group, 1977.

Dietrich, Thomas G., Freeman, Cecilia; and Crandall, Jo Ann. "A Linguistic Analysis of Some English Proficiency Tests," paper presented at the conference of the National Association of Bilingual Education, Seattle, Washington, May 1979.

El Paso Public Schools, Department of Curriculum & Staff Development. Oral Language Dominance Measure, Technical Information. El Paso, TX: March, 1978.

El Paso Public Schools, Department of Curriculum & Staff Development. Oral Language Proficiency Measure, Technical Information. El Paso, TX: Spring, 1977.

Herbert, C. H. Basic Inventory of Natural Language, Instructions Manual. San Bernadino, CA: Checpoint Systems.

Lau v. Nichols. 414 U. S. 563, 1974.

Office for Civil Rights, Department of Health, Education & Welfare. Task force findings specifying remedies available for eliminating past educational practices ruled unlawful under *Lau v. Nichols*. Summer, 1975.

Nie, Norman H. et al. Statistical Package for the Social Sciences. New York: McGraw-Hill, Second Edition Book Company, 1975.

Schutt, D. L. Schutt Primary Language Indicator Test, Examiner's Manual. New York: McGraw-Hill Book Company, 1976.

Texas Education Agency, Texas State Plan for Bilingual Education. Austin, TX: December, 1978.

Thurstone, L. L. Multiple-Factor Analysis. University of Chicago Press. Chicago ILL. 1947.